

Weekly Report

2017.06.05-2017.06.11

1. This Week

Wavelines

- 1.read the review of this submission and summarize its advices
- 2.according to the primary review's advice, we need to split this work into 2-3 papers. Think of what we can do once we split this paper.

(Details of the review advices and my thoughts are attached below)

Others

- 1.Revise the 5th chapter of the treatise: uncertainty data visualization
- 2.Translate the group work PPT to english version
- 3.Talk to every PHD student about their projects shown in the group work PPT and prepare for the introduction. Personally, I don't regard the presentation as a very successful one because we fail to control the entire presentation time and to distinguish between the detail and brief cases. For the next time, we need to emphasize the important points at the very beginning, combine the introduction of similar works and pay less attention to the very old work.

2. Review

得分: 2 (primary) 2 2.5 2.5 3

2.1 review总结

主要问题:

1. (写作上) 文章内容偏多, 在进行写作的时候没有区分主次、重要性、是否相关性
2. (写作与工作上) 文章的evaluation部分比较薄弱。从工作上来说, 对此系统的评估不够solid, 仅通过informal interview的形式得到了开放性的结果, 未实施打分评估体系, 且进行评估时未考虑到参与系统研发的专家应不能参与系统评估, 导致系统实用性的置信度不足; 从写作角度上来说, 整个evaluation部分不够balanced, 写作专家意见的同时要保证正面意见和负面意见的百分比, 突出正面意见, 通过负面意见对本文工作提供future work方向和改进方向。
3. (系统方面) 系统包含了非常多的元素、主题、应用方法, 但这些内容之间缺乏联系与逻辑, 使系统看上去缺乏整体性与连贯性

问题总结:

0.general

- 明确每个不同的theme是怎么fit到系统里面的, 并且他们互相之间的联系是怎么样的

- 哪些design是对于analyst来说最有用的，摒弃重要度低的设计，突出重要度高的设计及其中的技术方法

1.design tasks, requirements, design guideline

- 这些内容是否都与文章相关，他们怎么互相联系，并体现出与这个technical domain之间的紧密联系而不是从宽泛的角度提出需求或者设计指导。
- 文中的task analysis不合理，问专家他们需要做什么并不是task analysis，需要对任务和方法做进一步的抽象

2. experts

- 详细描述文中的experts，包括他们的数量、在实际工作中扮演的角色与工作内容、与他们互动的细节、他们平时是如何完成他们的工作的，有助于对此领域的理解
- 专家的建议如何与我们的技术决策关联起来（体现互动性）

3.evaluation

- Evaluation（case study）中用到的数据需要进行描述，以方便读者理解
- 介绍Case study是如何产生并与系统的哪方面内容相联系的
- Interview的结果要正反面都讲，并且参与系统研发的两个专家不能参与
- Evaluation需要增强系统的置信度，可以让用户对表格1中提出的任务进行打分(加强evaluation的方法与可信度)

4.color scheme

- 考虑到色盲患者，系统需要增加别的视觉认知通道，或者更改配色使其对色盲患者更为友好

5.related work

- Related work中的2.1 2.3部分没有应用到文章中来，是否考虑重构related work或者在系统中体现相关的概念

6.Others

- 系统目前在instability reasoning这方面较为薄弱，只能提出可能性但不能确认这种可能性，下一步工作中是否需要加强这部分内容
- 介绍清楚表现电网拓扑结构的Node-link图的scale-up方法

下一步工作计划：

6.12-6.25:针对此工作各个可深入研究的方向，对文章进行拆分，分析在每个方向上可以做的工作与思路。之后，对这些idea进行汇总、分析，选择近期应最先切入的方向开始对系统的修改与文章的修改。

6.25-7.30:修改系统，文章修改

8.1-8.15:润色文章、完善系统细节、录制视频。

难以解决的问题：

- 1.将专家意见与领域性强的问题与所选择的解决方案相联系，使技术选择具有指向性
- 2.如何针对专业性、领域性强的系统设计合理的、置信度高的evaluation流程

评语的启发：

- 1.之前在review论文的时候往往容易忽略所列的task, goal, requirement, design guideline与文章解决的问题或背景应用领域是如何联系的，因此这次在写作waveline的过程中对任务、需求、设计指南的抽象度不够，与背景领域的结合度也不够。
- 2.同样，以往在review文章的过程中往往只关注case是否有趣、是否体现出工作的意义，而往往忽略case与系统的联系，一个文章的不同case应该对应系统的需要突出的点，并且对应整个系统的逻辑连贯性，而不是简单的列举看上去有意思的案例。
- 3.文章的写作中需要有轻有重，学会在不同内容之间作取舍。一篇10页的论文可能无法介绍全系统的方方面面，在这种时候，写作过程中需要突出创新点、理念、逻辑关系，才能让评审抓住我们想表达的主要思想。
- 4.论文的review涉及文章的每个角落，从背景、任务、需求、工作现状，到方法技术、可视化设计、案例、评估，每一部分都会成为评审challenge的点。在第一次写作（一稿）的时候往往很难照顾到文章的方方面面，因此对文章的二次、甚至反复修改尤其重要。因此需要合理的安排时间，留下充裕的时间对文章、系统、视频进行polish。

2.2 IDEA LIST

论文可以拆分的三个方向及现有想法：

- 1) 专门研究SQC算法与可视化的结合，用于不同的数据集。
- 2) 描述一系列两个变量的不稳定性的可视化的方法；
- 3) 描述如何和专家互动，完成电网的不稳定性研究。

这个星期，主要对第一个方向进行了思考，思考的内容如下：（SQC算法与可视化的结合）

主要思路：

提炼一种通用的基于SQC（本质上是统计上的分布概率问题），与可视化结合的方法模型，可以适用于不同应用场景下的不同数据集。（Anomaly detection方向）

可用数据（多变量时序数据）：

1. 电网暂态稳定分析的仿真数据
2. 气象空气质量传感器数据

可以面向的问题：

问题1：supervisory control of evolutionary individuals

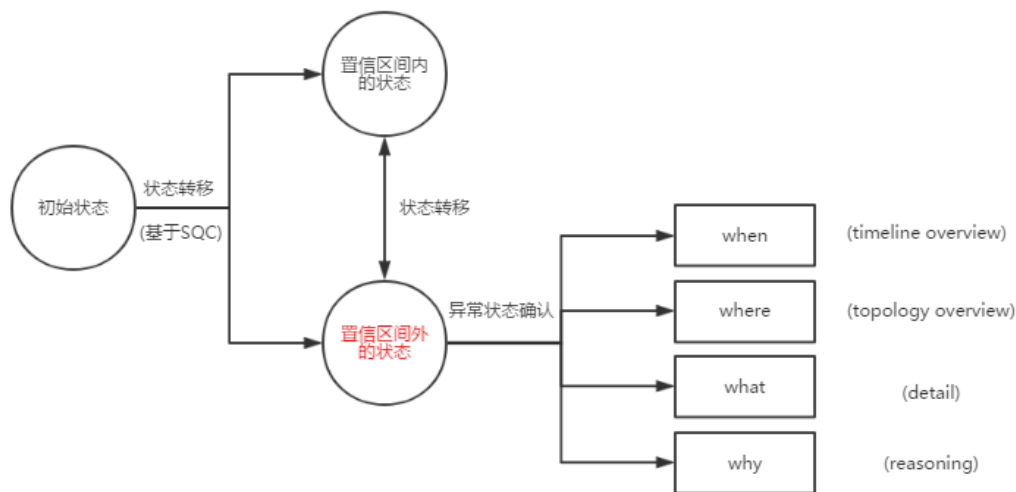
问题2：abnormal correlation analysis

以下想法主要针对解决问题1，有关问题2的内容下个星期将继续思考：

本质是对多变量时序数据中所有正在随时间发生变化的个体进行监督控制，利用基于SQC的可视化监测方法，定位相应应用场景中出现的异常模式/模式序列，解决应用场景中发生异常后的4W问题：when（何时），where（哪个/几个个体），what（出现什么样的异常模式/模式序列），why（为什么）。

体现在电网应用背景下，即：

现有的数据是电网暂态稳定分析的仿真数据，已知所施加的故障类型、施加故障的时刻、施加故障的设备，通过对电网设备的多变量时序变化模式来分析在相应的位置上施加相应故障后电网各部分的变化情况。那么，相反过程的研究是否同样具有意义。若仅已知电网设备的多变量时序变化情况，通过可视探索，故障发生的时刻、发生故障的设备、确定发生的故障类型。这就是一个监督控制的过程。



想法：

1. 定义一个网络在某个时刻（电网/传感器网络）的状态，考虑到每个节点都包含一个多变量组成的向量
2. 利用所有时刻的网络状态，计算网络状态的分布情况。分两种可能性：a），把高维的网络状态降维至一维后，就是普通的统计分布问题，利用SQC的检测概念，区分在指定置信区间外的异常状态；b），直接计算高维分布，利用SQC方法的变形，检测指定置信区间外异常状态（b方案涉及到如何将只能检验一种变量分布的统计检测方法用于多个变量的分布检测，需要进一步的思考）。
3. 对确定的异常状态进行4W可视分析

思考：

- 2015年vast的best paper将动态网络离散化为一个个snapshots并投影到二维空间，通过node-link图的聚类 and 演变发现动态网络演化阶段的各个stage。相似的，我们可以使用类似的方法展示系统演变的timeline和主要阶段，便于异常检测。

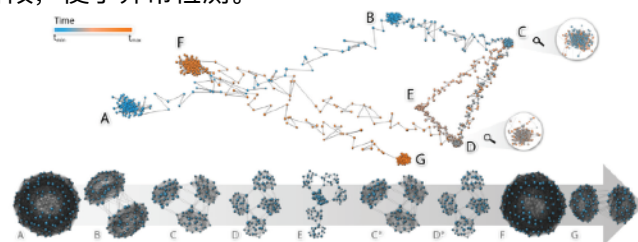


Fig. 1. Reducing snapshots of the dynamic network to points enables the exploration of network evolution. The projection of snapshots (top) reveals that the network contains 7 stable states (A-G), with 2 recurring states (C,D), and shows the transitions between them. A representative network for the snapshots in each stable state is shown (bottom).

- 利用统计方法进行异常检测，都需要对统计参数的选择即调整的讨论。不同于传统的参数调整的数值化面板，是否可以通过可视化的方法交互的调整统计方法中的参数。例如，用box plot进行参数调整，在展示数据分布的基础上交互的调整参数，是否有利于方法的理解？
- SQC的概念不仅可以用于正常状态与异常状态的区分，在具体分析4W问题的时候也能起到重要作用。之前的Waveline系统，利用SQC检测各个电网设备的某个变量值随时间变化的过程中异常的时间区段（when），类似的，我们是否可以考虑：纵向的去找电力系统中可能产生的问题的结构（where），指定一个特定的时刻，将这个时刻上所有个体的值（归一化后进行比较）可同样看做一个统计分布，对于各个时刻，去寻找各个分布中的离群异常点，即可确认发生问题的个体。
- 之前的椭圆图的设计保留并在此基础上继续发挥可视化的空间，做到可视化方面的创新。

3. TODO

1.treatise writing

2.Think about the next step of revising Wavelines (including splitting the paper and abstracting the models and methods)